

# Automatic detection of fatigue crack paths using digital image correlation and convolutional neural networks

Tobias Strohmann<sup>1</sup>  | Denis Starostin-Penner<sup>1</sup> | Eric Breitbarth<sup>1</sup>  |  
Guillermo Requena<sup>1,2</sup>

<sup>1</sup>German Aerospace Center (DLR),  
Institute of Materials Research, Cologne,  
Germany

<sup>2</sup>Metallic Structures and Materials  
Systems for Aerospace Engineering,  
RWTH Aachen University, Aachen,  
Germany

## Correspondence

Tobias Strohmann, German Aerospace  
Center (DLR), Institute of Materials  
Research, Linder Hoehe, 51147 Cologne,  
Germany.  
Email: tobias.strohmann@dlr.de

## Abstract

The occurrence of fatigue cracks is an inherent part of the design of engineering structures subjected to nonconstant loads. Thus, the accurate description of cracks in terms of location and evolution during service conditions is mandatory to fulfill safety-relevant criteria. In the present work, we implement a deep convolutional neural network to detect crack paths together with their crack tips based on displacement fields obtained using digital image correlation. To this purpose, fatigue crack propagation experiments were performed for AA2024-T3 rolled sheets using specimens with different geometries. Several hundred datasets were acquired by digital image correlation during the experiments. A part of the displacement data from one of the specimens was then used to train the neural network. The results show that the method can accurately detect the shape and evolution of the cracks in all specimens. Adding synthetic data generated by finite element analyses to the training step improved the accuracy for cracks with stress intensity factors that exceeded the range of the original training data.

## KEYWORDS

2024 T3 aluminum alloy, artificial neural network (ANN), crack lengths, fatigue crack growth, mechanical testing

## 1 | INTRODUCTION

Profound knowledge about fatigue crack propagation (*fc**p*) is safety-relevant for engineering materials subjected to nonconstant loads.<sup>1</sup> Its precise investigation is particularly relevant for airframe structures, for which the occurrence and growth of cracks are inherent to their design.<sup>2,3</sup>

In recent years, digital image correlation (DIC) has become a valuable instrument for enhanced data generation in experimental mechanics.<sup>4,5</sup> For instance, full field

displacement or strain data obtained by DIC are used in fracture mechanics<sup>6</sup> to determine stress intensity factors (SIFs),<sup>7</sup> J-Integral,<sup>8</sup> or local damage mechanisms.<sup>9,10</sup> Investigations of crack propagation or local crack tip fields by DIC require accurate information about the crack path and especially the crack tip position. Image-based methods such as the Sobel edge-finding routine can be applied to identify crack paths because the open crack leads to a large displacement field gradient in its surroundings.<sup>11</sup> Moreover, the characteristic crack tip strain field can be utilized to determine the actual crack

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Fatigue & Fracture of Engineering Materials & Structures published by John Wiley & Sons Ltd.

tip coordinates by fitting the Williams' series expansion<sup>12</sup> to the experimental data set.<sup>13,14</sup> As a side effect, the crack tip loadings as well as higher-order terms are also determined by this method.<sup>15</sup> In contrast to classical crack length methods like direct current potential drop (DCPD)<sup>16,17</sup> or the compliance method,<sup>18,19</sup> image analysis techniques are able to detect cracks with various shapes.<sup>20,21</sup> However, a fully automatic detection of the crack path, and especially the crack tip, is often limited owing to experimental scattering or artifacts in the DIC dataset.<sup>22,23</sup> Thus, the application of DIC for a large number (several hundred) of images during an *fc*p experiment always comes along with extensive manual work which constitutes a bottleneck in the research process.

Machine learning and, more specifically, deep learning are becoming promising tools for structural health monitoring in civil engineering. This involves in particular surface inspection to detect cracks in buildings,<sup>24,25</sup> roads,<sup>26,27</sup> or metallic structures.<sup>28,29</sup> In these methods, the inspection is done using 2D images of the damaged structure. The issues addressed in these studies are basically very similar to fracture mechanics problems, although they do not require a high accuracy regarding the detection of the crack tip position.

In the present work, we trained a convolutional neural network (CNN) for the detection and segmentation of crack paths and crack tip coordinates in physical data (i.e., the displacement field) obtained from DIC measurements during an *fc*p experiment. First, we generated manually labeled data, which were used for training and validation of the CNN. Secondly, we trained the network and analyzed its performance with a second, thicker specimen. Finally, we applied the CNN to a third DIC data set of a much larger specimen. We analyzed the accuracy of the segmented crack tip position and the resulting a-N curves to quantify the performance of the CNN. Furthermore, we investigated the effect of additional training data obtained by finite element analysis (FEA) to complement experimental data.

## 2 | METHODOLOGY

### 2.1 | Material

The commercially available AA2024-T3 aluminum alloy was used for the *fc*p experiments. This alloy was chosen because it represents one of the most relevant alloys applied in the aircraft industry, particularly for fuselage structures.

Middle tension (MT) specimens were cut from rolled sheets with thicknesses of 2.0 and 4.7 mm.

A linear-elastic material model was used for the corresponding finite element simulation with a Young's modulus  $E = 72.0$  GPa and Poisson's ratio  $\mu = 0.3$ .

### 2.2 | Experimental setup and image acquisition

The experimental setup follows the standard testing procedure for *fc*p tests according to ASTM E647.<sup>30</sup> A servo-hydraulic testing rig with a maximum force capacity of 60 kN was used. The MT specimens had a width  $W = 160$  mm with an artificial starter notch of  $2a \approx 20$  mm in its center introduced by a saw cut. One specimen with thickness  $t = 4.7$  mm (specimen  $S_{4.7}$ ) and one with thickness  $t = 2.0$  mm (specimen  $S_{2.0}$ ) were tested applying a load ratio  $R = 0.1$ . Specimen  $S_{4.7}$  and  $S_{2.0}$  were loaded with  $\Delta F = 45.0$  kN and  $\Delta F = 13.5$  kN, respectively. Nominal stresses of 60 and 42 MPa for  $S_{4.7}$  and  $S_{2.0}$ , respectively, resulted in different evolutions of the SIF  $K_I$ . The initial crack lengths  $a_{\text{init}}$  and the maximum crack lengths at final rupture  $a_{\text{rupture}}$  are given in Table 1 together with their corresponding SIF. Taking these conditions into account, the experiments cover a SIF range from  $\sim 1$  to 35 MPa $\sqrt{\text{m}}$ .

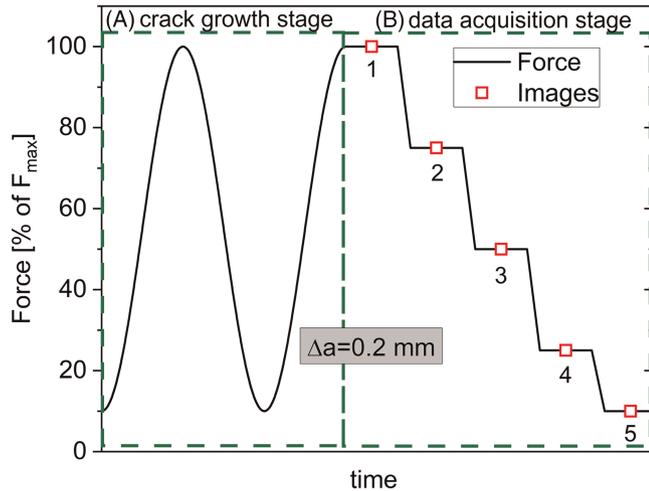
A Canon 70D reflex camera with a 90 mm macro lens was positioned at the specimen's front side, while the commercial GOM Aramis 12 M 3D DIC system was positioned at the back side. The DIC system has two cameras with a lens focal length of 50 mm that were placed 240 mm apart from each other and at a distance of  $\sim 630$  mm from the sample. A spatial resolution of  $\sim 1.2$  mm (1 pixel  $\sim 0.08$  mm) with a facet size of  $19 \times 19$  pixels and a facet distance of 15 pixels was used for DIC. The front side of the sample was polished to enhance the visibility of the crack path. The back side was covered with a stochastic black/white pattern used for the DIC calculations.

Two different stages were subsequently repeated during the experiment:

1. Crack growth stage (Figure 1A): The cyclic force  $\Delta F$  was applied with a frequency of 15 Hz, and the crack length was determined by the DCPD method using Johnson's equation<sup>31</sup> to control the experiment.
2. Data acquisition stage (Figure 1B): The experiment was stopped for a few seconds at maximum load every 0.2 mm of crack extension  $\Delta a$ . During that time, an image of both sides of the specimen was taken simultaneously, that is, one with the DIC system and one with the reflex camera. The force was then reduced in

**TABLE 1** Crack lengths and load conditions of specimens  $S_{4.7}$  and  $S_{2.0}$ 

	Thickness $t$ (mm)	$2a_{\text{init}}$ (mm)	$2a_{\text{rupture}}$ (mm)	$K_{\text{init,min}}$ (MPa $\sqrt{\text{m}}$ )	$K_{\text{rupture,max}}$ (MPa $\sqrt{\text{m}}$ )
$S_{4.7}$	4.7	21.5	100	1.2	35
$S_{2.0}$	2.0	17.4	120	0.8	33

**FIGURE 1** (A) Cyclic loading of the specimen with a force ratio  $R = 0.1$  during crack growth stage. (B) The data acquisition stage begins at crack length increment  $\Delta a \geq 0.2$  mm [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

four successive steps (75%, 50%, 25%, and 10% of the maximum) to acquire five images at different load levels at constant crack length.

### 2.3 | Data preparation

The DIC data sets are stored as a list of displacement vectors  $u$  with  $x$  and  $y$  components (i.e.,  $u_x$ ,  $u_y$ ) measured with respect to the coordinate of each facet's center point (see Figure 2A). The  $z$  component  $u_z$  was neglected because displacements in the thickness direction were expected to be comparably small. The coordinates are not necessarily structured in a perfect grid with respect to the macroscopic Cartesian coordinate system. In addition, the data sets can have missing entries owing to failed calculations of the facet data. Nevertheless, a regular grid of  $256 \times 256$  entries (Figure 2C) covering a field of view of  $70 \times 70$  mm<sup>2</sup> was used as input for the CNN. Therefore, the DIC data (i.e., displacements  $u_x$ ,  $u_y$ ) were interpolated from the facets' center points onto this  $256 \times 256$  matrix (Figure 2B). This procedure ensures an equally sized input and target grid for the CNN with an interpolation point distance of  $\sim 0.27$  mm. The size of  $256 \times 256$  was chosen as a good trade-off between computational effort

and resolution (i.e., higher resolution than the DIC system; therefore, we are not limiting the spatial resolution by the interpolation).

### 2.4 | Implementation of the neural network

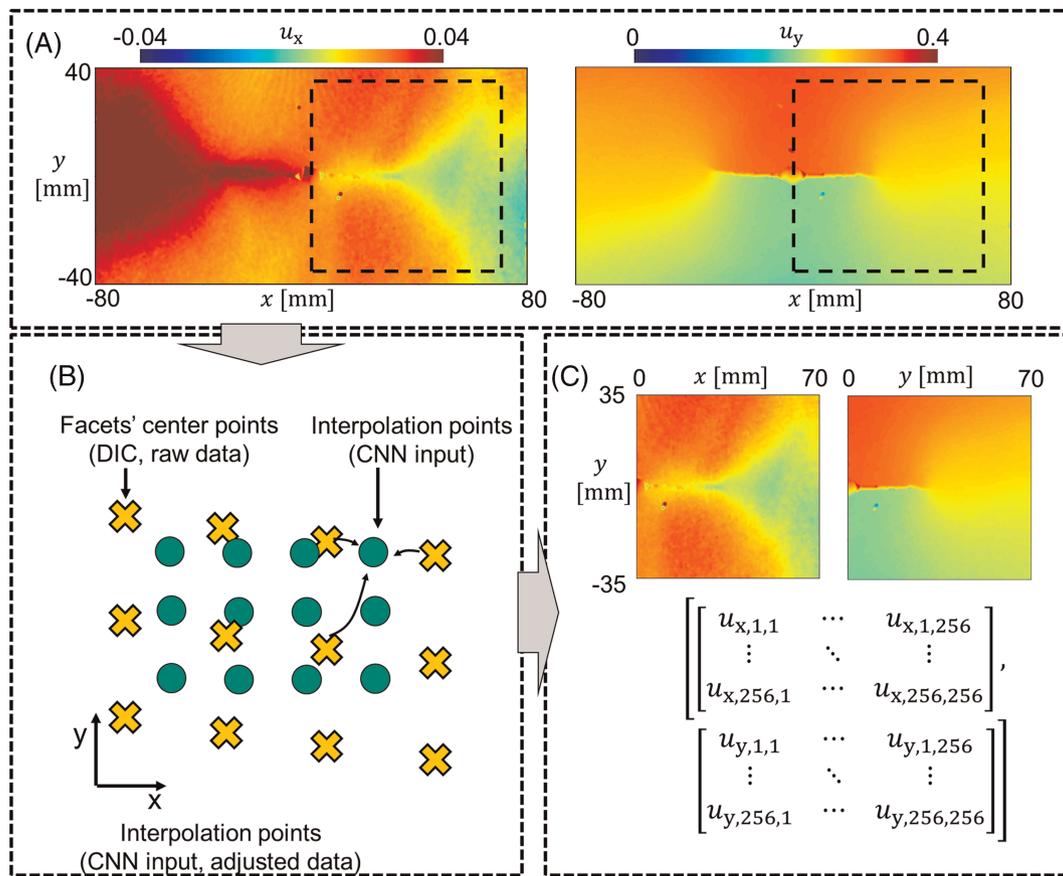
A CNN with a U-Net architecture, introduced by Ronneberger et al.,<sup>32</sup> with a handful of changes achieved from trial-and-error, was used in the present work. U-Net is a state of the art CNN architecture used for image segmentation tasks in a broad range of fields. In literature, different architectures are available for semantic segmentation (e.g., Fully CNNs,<sup>33</sup> ParseNet,<sup>34</sup> and later approaches like Mask R-CNN<sup>35</sup>). However, we expect that our use case, that is, the detection of a crack in DIC data depends primarily on the data used for training. The U-Net architecture has a relatively low complexity and is easy to adept by trial and error. Furthermore, it has been successfully used for detecting cracks in images of concrete.<sup>36</sup> The architecture consists of an encoding and a decoding part with additional skip connections that transfer information from the encoder to the decoder to increase the segmentation accuracy.<sup>37</sup> A summary of the implemented architecture is given in the supplementary Figure S1 and in supplementary material S2. We used a *Leaky ReLU* function as activation function between the convolutions instead of the usually applied *ReLU*.<sup>38</sup> This was done because *ReLU* sets negative values to zero:

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases}$$

Instead, *Leaky ReLU* has a small gradient if  $x \leq 0$ , which seems reasonable for physical data that may contain negative values, as it is the case of the displacement fields used in this work:

$$\text{Leaky ReLU} = \begin{cases} x, & \text{if } x > 0 \\ 0.01x, & \text{if } x \leq 0 \end{cases}$$

The implementation of the CNN was done in Python 3 using the machine learning framework PyTorch<sup>39</sup> and standard libraries for data processing. The Adam



**FIGURE 2** (A) Graphical visualization of the digital image correlation (DIC) measurements: The colors indicate the horizontal and vertical displacements ( $u_x$ ,  $u_y$ ) with respect to each facet's center point. (B) The displacements with respect to each facet's center point are interpolated on a  $70 \times 70$  mm<sup>2</sup> interpolation grid, with  $256 \times 256$  grid points. (C) Visualization of the regular array with a resulting size of  $2 \times 256 \times 256$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

optimizer<sup>40</sup> (learning rate =  $1 \times 10^{-4}$ , amsgrad = True) and PyTorch's cross entropy loss function were used for training the network.

The horizontal ( $u_x$ ) and vertical ( $u_y$ ) components of the displacement field acquired by the DIC system were used as input channels. The output channels of the network are constituted by three classes: (1) background, (2) crack path, and (3) crack tip. If the segmentation works properly, the crack tip should be identified as a single pixel, the crack path should have a length of a few hundred pixels (single line), and the background size should be  $\sim 65 \times 10^3$  pixels ( $256 \times 256$  pixels minus crack tip and crack path). Thus, the three output classes are drastically imbalanced in terms of number of pixels. In such a case, the training process must be supported by adding weight factors for each class when calculating the loss function.<sup>41</sup> Therefore, different weight factors were tested, resulting in the chosen weights  $\omega_{BG} = 1$ ,  $\omega_{CP} = 350$  and  $\omega_{CT} = 500$  for the classes (1) to (3), respectively. The network was trained with a batch size of 25 inputs per batch.

Training and testing of the CNN were performed on a Fujitsu Celsius R930 Power Work Station. The machine has two Intel Xeon E5-2667v2 3.30 GHz 25 MB Turbo Boost processors with eight kernels each and is equipped with  $12 \times 16$  GB DDR3 working memory. An NVIDIA RTX 8000 graphics card was used for GPU acceleration.

## 2.5 | Generation of ground truth data

The generation of ground truth data is one of the most challenging parts in any supervised machine learning application. The data must contain an appropriate number of training samples which are also sufficiently error-free.

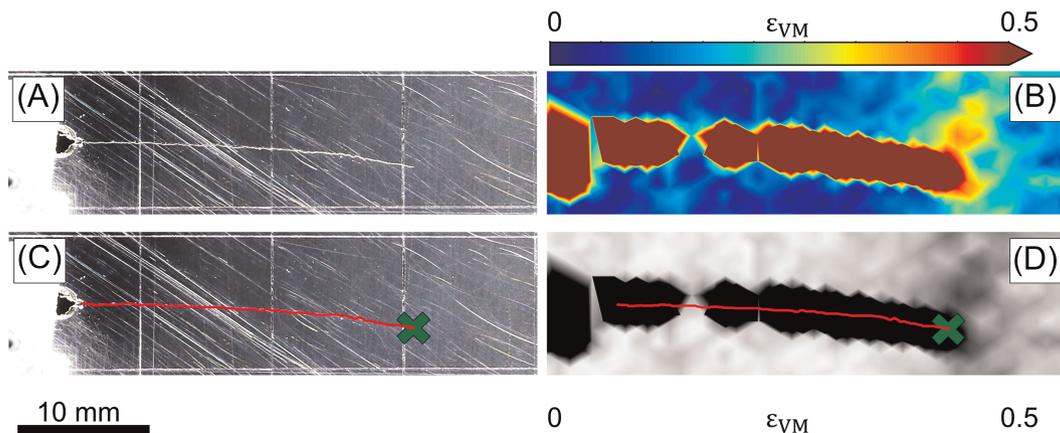
Ground truth data were obtained by manual segmentation of the crack tip and the crack path from 160 optical images with different crack lengths. These images were acquired with the digital reflex camera during the *fc*p experiment. Representative images of the segmented

crack path and crack tip are shown in Figure 3A (before segmentation) and Figure 3C. The crack path and crack tip were assumed to be identical for both specimen sides, that is, front side (reflex camera) and back side (DIC system). Figure 3B,D shows the corresponding von Mises strain field obtained by DIC as well as the superimposed manually segmented crack path and crack tip, respectively. The same path and crack tip coordinates were used as ground truth for the five force steps of the acquisition stage (see Figure 1B).

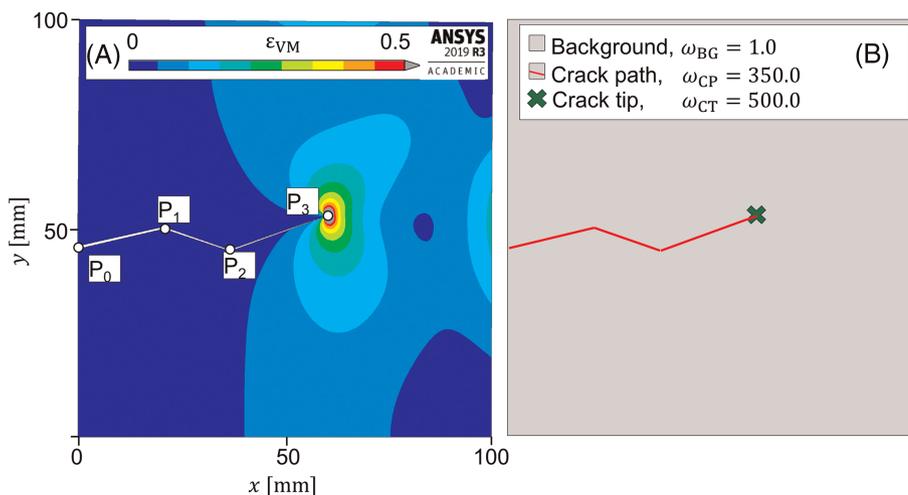
It is worth mentioning that the specimen showed a  $45^\circ$  sheared fracture surface (S-mode) at crack lengths  $a > 30\text{--}40$  mm. Thus, the coordinates of the crack paths on the front and back side of the specimen were not identical. Although a manual correction that takes into account the angle of the fracture surface was carried out, the segmentation deviation between both sample sides is expected to be larger than for a flat crack.

### 2.5.1 | Generation of ground truth data using FEA

FEA data were included in the training process additionally to experimental DIC data. To this purpose, a crack with a horizontal orientation and random deviations along its path was introduced in a  $100 \times 100$  mm<sup>2</sup> 2D model with plane elements (see Figure 4A). Mean  $K_I$  of the FEA data was  $11.5$  MPa $\sqrt{\text{m}}$ , with a minimum and maximum  $K_{I,\text{min}} = 2.32$  MPa $\sqrt{\text{m}}$  and  $K_{I,\text{max}} = 33.49$  MPa $\sqrt{\text{m}}$ , respectively. This load range was chosen similar to the experiments. Random coordinates of the points  $P_0$  to  $P_3$  determine the crack path, where  $x_0 = 0.0 < x_1 < x_2 < x_3$  and  $y_0 < y_1 > y_2 < y_3$ . Crack deviation angles were  $< 22.5^\circ$  with respect to the horizontal axis. Therefore, crack length was equally distributed between 50 and 72 mm. Maximum  $y$ -offset of the crack tip was  $\pm 42.5$  mm with respect to the center of the specimen. The elastic material properties given in Section 2.1



**FIGURE 3** Generation of ground truth data: (A) macrograph of the crack (half length) obtained with the reflex camera. (B) Von Mises equivalent strain field,  $\epsilon_{VM}$ , obtained on the back side of the specimen by digital image correlation (DIC). (C) Manually segmented crack path and crack tip. (D) Manually segmented crack path and crack tip superimposed to the DIC data [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 4** (A) Exemplary von Mises strain ( $\epsilon_{VM}$ ) field obtained by finite element analysis (FEA). The positions  $P_0$ – $P_3$  are randomly generated to create the crack path. (B) Resulting ground truth image. The imbalance between the classes *background*, *crack path*, and *crack tip* in relation to the area fraction can be seen in this figure [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

were used for calculations under 2D plane stress conditions. The 2D linear–elastic calculations are a simplification of a fatigue crack in a ductile material. However, the characteristic crack tip field is a good approximation, since the plastic region is very small compared to the crack length. The sample was loaded vertically, while the vertical and horizontal displacements were again interpolated within a  $256 \times 256$  entries matrix, analogously to the data from DIC measurements. The advantages of using FEA data for CNN training are as follows:

1. The generation of FEA-based ground truth data is extremely time and cost efficient compared to experimental DIC data (see Figure 4A).
2. The diversity of samples (crack paths) and loading conditions can be easily increased.
3. The ground truth is generated automatically and has no bias from manual labeling (see Figure 4B).
4. FEA reflects accurately the theoretical stress field in the vicinity of the crack tip without the effect of experimental scatter (see Figure 4A).

## 2.6 | Data augmentation

Usually, data augmentation is applied to increase the number of training images as well as the diversity of training data by translation, rotation, stretching, zooming, and color manipulations of the original images. However, most of these operations cannot be applied on data that contain actual physical information as it is the case of the displacement fields around a crack tip in the present work. Therefore, the five different load steps (see Figure 1) were used here as a sort of data augmentation, since these images contain different magnitudes of  $u_x$  and  $u_y$  displacement that correspond in fact to the same condition, that is, the crack path does not change. Additionally, all images from the left half of the crack were mirrored to the right side. Thus, only one crack growth direction, from left to right,

was used. Furthermore, each crack was mirrored vertically to double the number of samples for the training. Also, the position of the interpolation array (Figure 3A) was shifted vertically in each image to vary the position of the crack with respect to the interpolation array. This was done randomly with maximum offset values of  $\pm 17.5$  mm ( $= 1/4$  height of the field of view).

## 2.7 | Data sets and data splitting

A total of 1668 and 2838 samples of input and ground truth were obtained after augmentation of the DIC data for  $S_{4.7}$  and  $S_{2.0}$ , respectively. Additionally, 2668 samples were generated by FEA.

Several DIC measurements were acquired under stress-free conditions (i.e., without external load) for  $S_{4.7}$  before starting the fatigue crack growth. Based on that, an additional DIC data set was computed using a different DIC reference, that is, using a second image acquired at zero external load. We used this second data set to quantify the impact of background noise of the DIC system on the performance of the CNN.

The data were split into six different subsets (see Table 2):

1. Training data set  $train_{4.7}$ : 80% of the DIC data of  $S_{4.7}$  doubled by vertical mirroring, resulting in  $1668 \times 0.8 \times 2 = 2668$  samples.
2. Training data set  $train_{S_{4.7} + FEA}$ : The DIC data of  $train_{S_{4.7}}$  plus the same amount of FEA calculations, resulting in  $2668 \times 2 = 5336$  samples.
3. Test data set  $test_{S_{4.7}}$ : 20% of the DIC data of  $S_{4.7}$ , that is,  $1668 \times 0.2 = 333$  samples.
4. Test data set  $test_{S_{4.7}, Ref2}$ : 100% of the DIC data of  $S_{4.7}$  but recomputed with a DIC reference different from the one used for  $train_{DIC}$ .
5. Test data set  $test_{S_{2.0}}$ : 100% of the DIC data of  $S_{2.0}$ .
6. Test data set  $test_{FEA}$ : consisting of 333 samples from the FEA calculations.

**TABLE 2** Definition of data sets used for training and testing

Training/test	Name	DIC $S_{4.7}$	DIC $S_{2.0}$	Data from FEA
Training	$train_{S_{4.7}}$	2668	0	0
Training	$train_{S_{4.7} + FEA}$	2668	0	2668
Test	$test_{S_{4.7}}$	333	0	0
Test	$test_{S_{4.7}, Ref2}$	1668 (new reference stage)	0	0
Test	$test_{S_{2.0}}$	0	2638	0
Test	$test_{FEA}$	0	0	333

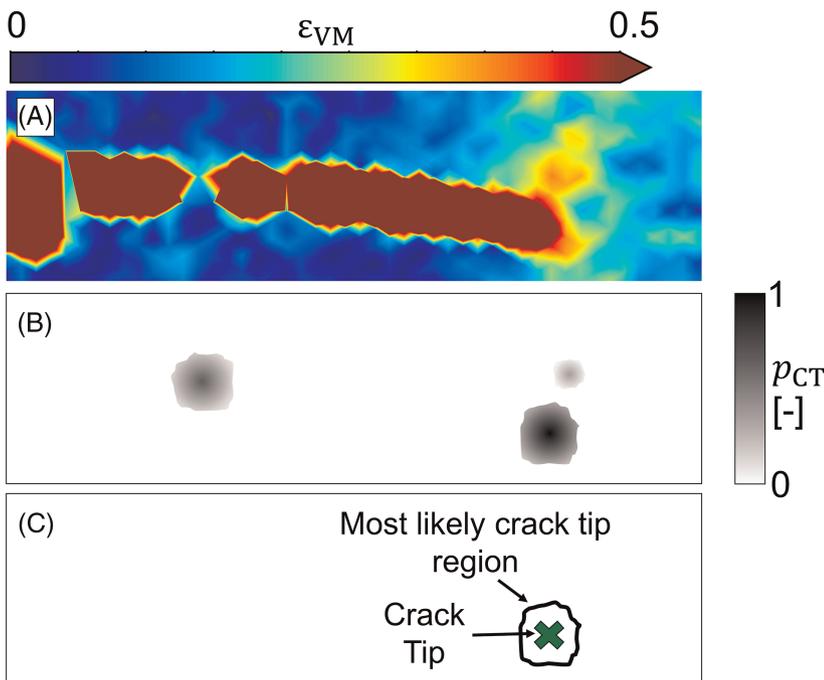
Abbreviations: DIC, digital image correlation; FEA, finite element analysis.

## 2.8 | Post-processing

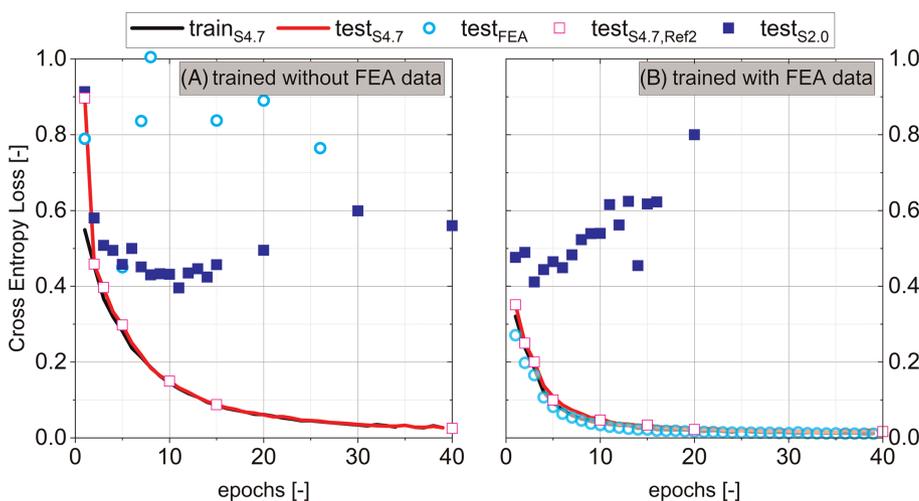
Generally, the procedure of semantic segmentation consists of the following steps:

1. The output of the neural network is generated with respect to the number of classes (here: *background*, *crack path*, and *crack tip*).
2. The output is transformed into probabilities,  $p_i$ , for each class  $i$ , where  $\sum_i p_i = 1$ .
3. A pixel/entry of the matrix is classified based on the highest probability among the possible classes (max voting).

However, this scheme may result in uninterpretable results for the present case because several pixels may become classified as *crack tip* although this class' ground truth consists in only one pixel. A post-processing routine was applied to tackle this problem (see example in Figure 5). The algorithm checks first if multiple regions of the results array are classified as *crack tip* (Figure 5B). If more than one region is found, then the one with the highest mean probability of class *crack tip*,  $p_{CT}$ , is chosen (Figure 5C). Finally, the center of gravity of this area is defined as the resulting crack tip coordinate. This procedure is only applied for the test case and has no effect on the backpropagation of loss values.



**FIGURE 5** (A) Example of the von Mises equivalent strain measured by digital image correlation (DIC) around a crack. (B) The convolutional neural network (CNN) results array may contain multiple regions with a high probability for class *crack tip*. (C) Only that region with the highest mean probability is chosen. Finally, the algorithm indicates the gravity center of that region as *crack tip* [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 6** Evolution of the cross entropy loss as a function of trained epochs. (A) Training and test loss functions without finite element analysis (FEA) data during training and (B) training and test loss functions with additional FEA data during training [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 3 | RESULTS AND DISCUSSION

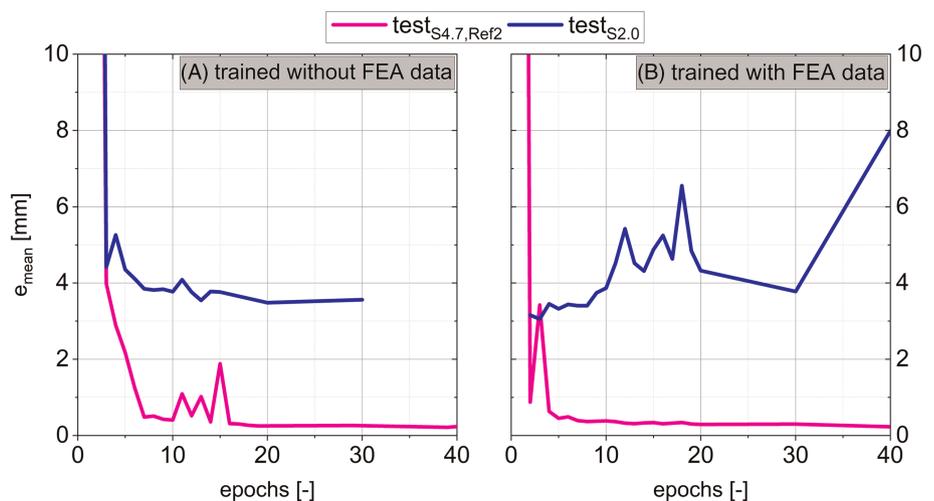
Figure 6A,B shows a convergence of the cross entropy loss function with increasing number of epochs for the CNN trained without and with FEA data, respectively. In both cases, the test with data set  $\text{test}_{S_{4.7}}$ , that is, from the specimen used for training, reveals a trend of the loss function similar to the training stage. Also, the test of the complete data set of  $S_{4.7}$  calculated with a different DIC reference ( $\text{test}_{S_{4.7},\text{Ref}2}$ ) shows a very similar trend. In each of these three cases, the training with additional FEA data results in a convergence after fewer epochs owing to the larger size of the data set (see Table 2). Additional FEA data during training increase the performance of the network for the pure FEA data set during testing (turquoise circles in Figure 6B). The data set  $\text{test}_{S_{2.0}}$ , that is, from the second sample, results in relatively high loss values in comparison to the  $S_{4.7}$  results. Moreover, a minimum in the loss function can be found after 11 and 3 epochs for the training without and with FEA data, respectively. Beyond these points, the network seems to overfit, resulting in an increase of the cross entropy loss.

Generally, the loss function is a good indication of the success of a training stage and gives an impression about the accuracy of the segmentation results. However, the three classes analyzed in this work have different degrees of importance from a fracture mechanics point of view, that is, background = low importance, crack path = medium importance, and crack tip = high importance. Thus, instead of classical metrics for semantic image segmentation (e.g., pixel accuracy, dice coefficient, or Jaccard index), a metric that takes into account the high importance of the crack tip position is necessary. Therefore, the mean mismatch in the identification of the crack tip position,  $e$ , was evaluated to quantify the performance of the trained CNN:

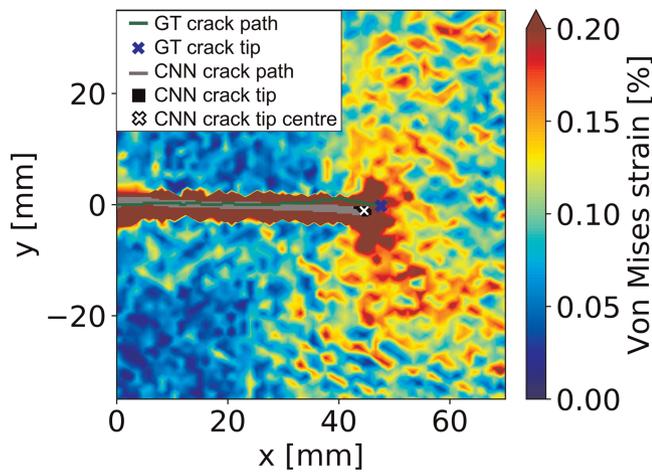
$$e = \sqrt{(x_{\text{GT}} - x_{\text{CNN}})^2 + (y_{\text{GT}} - y_{\text{CNN}})^2},$$

where the subscript GT indicates the ground truth position and CNN corresponds to the position predicted by the trained CNN. The results are shown in Figure 7A,B for training without and with FEA data, respectively. The figure represents the variations in the position of the crack tip for all crack lengths during the *fc*p test, which does not reflect the actual position of a crack tip at an individual crack length. In both cases, the mismatch  $e_{\text{mean}}$  for  $\text{test}_{S_{4.7},\text{Ref}2}$  is  $<0.5$  mm after a sufficient number of epochs ( $\sim 20$ ), showing spikes between epochs 10–16 for the training without FE data. This means that the performance of the trained network is independent from random scatter of DIC data between two stages with the same loading condition. The mismatch for  $\text{test}_{S_{2.0}}$  stabilized at 3–4 mm after  $\sim 10$  epochs for the training without FEA data, while it seems to become unstable for the training with FEA data after reaching a minimum at 3–4 mm between 3 and 10 epochs. While the mismatch  $e_{\text{mean}}$  for sample  $S_{2.0}$  is clearly larger than for  $S_{4.7}$ , its value is still reasonably small at 20 epochs (without FEA data) and 5 epochs (with FEA data). Moreover, these points are in the stabilized region of  $e_{\text{mean}}$  (Figure 7) and show a relatively low cross entropy loss (Figure 6). Therefore, we decided to use the network trained at these conditions for further analysis.

Figure 8 shows a representative image of the ground truth (green line with blue marker) and the predicted crack (gray line with black and white markers) for  $S_{2.0}$  superimposed to the von Mises equivalent strain determined by DIC. Taking into account the shape of the local crack tip strain field (von Mises strain), it seems probable that the CNN prediction (epoch 5, with FEA data) is even more accurate than the manually segmented ground truth based on the image from the other side of the



**FIGURE 7** Mismatch between manually segmented crack tip position and convolutional neural network (CNN) prediction (A) without additional finite element analysis (FEA) data and (B) with additional FEA data during training [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



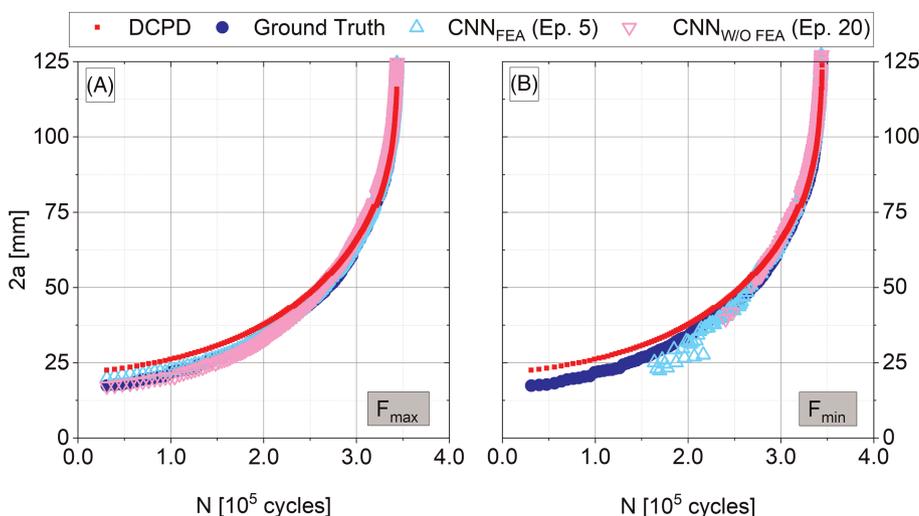
**FIGURE 8** Manual and convolutional neural network (CNN) segmentation of the crack path and the crack tip of  $S_{2.0}$ . Background showing the von Mises strain  $\varepsilon_{VM}$ . In this case, the CNN was trained with additional finite element analysis (FEA) data for five epochs [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

specimen. These results show that the network successfully indicates the crack path, including the crack tip. Furthermore, the results in Figure 8 show that the relatively large errors observed in the CNN assessment of specimen  $S_{2.0}$  (evolution of cross entropy loss and  $e_{\text{mean}}$ —Figures 6 and 7) give still reasonably good results. This apparent discrepancy may be understood considering, on the one hand, that the unstable evolution of the loss function after reaching a minimum (Figure 6) is caused by overfitting of the network to the training data. On the other hand,  $e_{\text{mean}}$  (Figure 7) is a function of the manual segmentation used as ground truth, which is biased by the ground truth generation procedure (see Section 2.5). In particular case of  $S_{2.0}$ , the illumination of the sample during the test was not optimal to

univocally identify the crack tip position. Nevertheless, the CNN segmentation in Figure 8 and more impressively the supplementary video Animation S3 clearly show a segmentation of the crack path and the crack tip position that seem to be more accurate than the manual segmentation based on images obtained from the reflex camera.

The evolution of the crack length,  $2a$ , as a function of the number of cycles,  $N$ , is shown for sample  $S_{2.0}$  in Figure 9 at maximum (A) and minimum load (B). There is a discrepancy between the DCPD curves and the ground truth because the DCPD curves were not subsequently corrected, that is, no markers were generated on the fracture surface (to keep the stress ratio  $R$  constant during the experiment). The CNN segmentations with and without additional FEA training data show a very good agreement with the ground truth segmentation at maximum load (Figure 9A). Furthermore, the crack growth curves evolve steadily without any unrealistic behavior confirming the robustness of the method. The supplementary video (Animation S3) shows the evolution of the crack path and crack tip segmented by the CNN trained with FEA data superimposed to the DIC von Mises strain field at maximum load. The video shows that the crack tip is usually detected inside the highly strained region. Furthermore, the crack path evolution is segmented accurately for all different crack lengths although the individual input arrays were evaluated independently from each other.

Crack lengths  $2a < 50$  mm show a larger deviation with respect to the ground truth for the cracks predicted by the CNN at minimum load (Figure 9B). Here, the crack tip coordinates were not detected for the training without FEA data or they were identified far from the correct position for the training with FEA data. These results are due to the fact that the shorter crack lengths resulted in smaller strains, which are more affected by



**FIGURE 9** Crack length  $2a$  for  $S_{2.0}$  as a function of cycles  $N$ : (A) at maximum applied force  $F_{\text{max}}$  and (B) at minimum applied force  $F_{\text{min}}$  (see Figure 1) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

background noise and scattering during the DIC calculations. This indicates that background noise and scattering have a strong influence at low applied loads on the specific image patterns that the network is trained to find in the displacement field data.

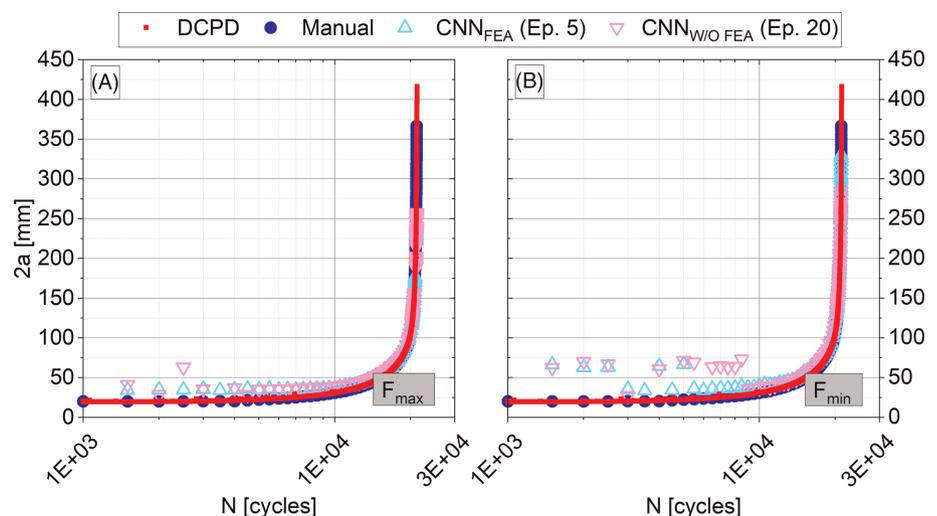
### 3.1 | Application of the trained CNN to a larger specimen

In the next step, the CNN is applied to a different specimen type. We used a large AA2024-T3 MT specimen with a width  $W = 950$  mm and a thickness  $t = 1.6$  mm. Fatigue crack growth was investigated under very high load conditions  $K_{I,max} \sim 30\text{--}140$  MPa $\sqrt{\text{m}}$ , corresponding to crack lengths  $25 < 2a < 370$  mm and a maximum nominal stress  $\sigma_{max} = 120$  MPa. The load ratio was  $R = 0.3$ . In contrast to the previous data set, the spatial resolution of the DIC system was about 4.8 mm (1 pixel = 0.32 mm) with a facet size of  $19 \times 19$  pixels and a facet distance of 15 pixels. Further details of the *fcg* experiment can be found in Breitarth et al.<sup>42</sup> A new training data set that includes the existing training data  $\text{train}_{S_{4.7}}$  and new FEA training data was used. The FEA simulations took into account the higher load range applied to this large specimen. Thus,  $K_I$  covered a range of 40–200 MPa $\sqrt{\text{m}}$  for FEA. For the CNN test we chose epoch 20 for training without FEA data and epoch 5 for training including the new FEA data. Furthermore, the size of the interpolation array (see Figure 2) was changed to  $250 \times 250$  mm<sup>2</sup>.

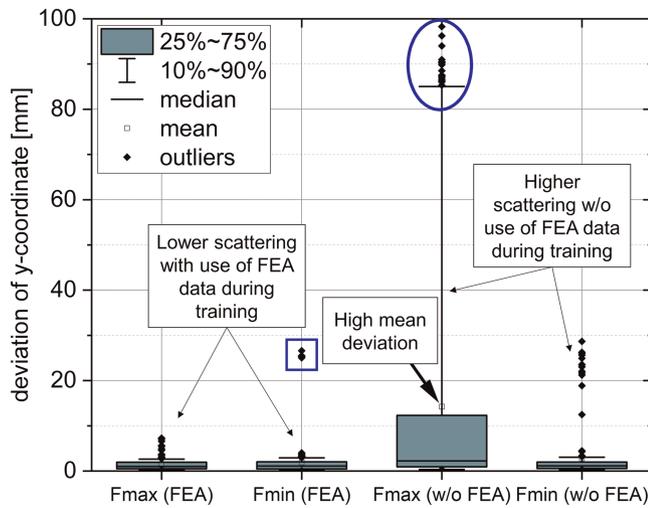
The results obtained by CNN segmentation with and without FEA data during training are shown in the  $a$ - $N$  diagram in Figure 10A,B for maximum and minimum load, respectively. Manually segmented crack lengths, obtained by rough estimations based on the equivalent

von Mises strain field, are also included in this figure. The CNN achieved an accurate segmentation for crack lengths  $40 < 2a < 250$  mm for both loading conditions. Moreover, the CNN segmentation at maximum load was also successful for shorter crack lengths (Figure 10A), while a precise prediction was achieved for longer cracks ( $2a > 250$  mm) at minimum load (Figure 10B). As described for specimens  $S_{4.7}$  and  $S_{2.0}$ , the shorter cracks result in small strains which are shadowed by the background noise of the DIC evaluation. Thus, the maximum load results in a better segmentation performance for crack length  $2a < 40$  mm. On the other hand, the strains around the crack at maximum load are much higher than for  $S_{4.7}$  owing to the higher  $K_I$  in this large specimen. This effect is less pronounced under minimum load, and therefore, longer cracks can be segmented more accurately in this condition. Including FEA data in the training process extends the range at which crack lengths are segmented correctly.

Generally, the  $y$  coordinate of the crack tip (i.e., the coordinate perpendicular to the crack path) is segmented more accurately by the CNN and it is also easier to identify during manual segmentation than the  $x$  coordinate. The deviation of the CNN prediction with respect to the manual segmentation is shown as a box chart in Figure 11. A strong impact of the use of additional FEA data during training can be seen. The best results are obtained with the use of FEA data at maximum load. There are only a few outliers with comparable low deviations ( $< 8$  mm) related to the manual segmentation. Similar results are obtained under minimum load, although some data points show deviations  $> 20$  mm (blue square). However, the median is located at 1.4 and 1.1 mm under maximum and minimum loads, respectively. The scattering of data is much more significant if no FEA data were used during training. More than 10% of data points show



**FIGURE 10** Results for a large middle tension (MT) specimen with a width  $W = 950$  mm. Manual and convolutional neural network (CNN) segmentation of cracks: (A) under maximum applied force  $F_{max}$  and (B) under minimum applied force  $F_{min}$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 11** Y coordinate deviation (i.e., perpendicular to crack path) for the 950 mm wide middle tension (MT) specimen after convolutional neural network (CNN) training with and without finite element analysis (FEA) data [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

a deviation  $>80$  mm (blue circle) under maximum load, which also results in a very high mean deviation  $>10$  mm (arrow). Those data points usually corresponded to very large crack lengths, analogously to the  $x$  coordinate (see Figure 10). The median is located at 2.2 and 1.1 mm under maximum and minimum loads, respectively.

It has to be mentioned that the outliers are usually a product of the post-processing routine. Although the CNN does find a crack tip close to the correct position, a second region of class *crack tip* is located far from the correct one (compare Figure 5). The correct region may be chosen mistakenly since only the region with the highest probability is chosen to generate a single crack tip (one pixel). The wrong region of class *crack tip* may be a result of artifacts in the DIC measurement which are, for example, caused by local defects of the spray paint that lead to unrealistically high displacements. A future version of the CNN implementation may tackle this issue by creating first a “possible crack tip region” window based on physical knowledge (e.g., displacement gradients) followed by the segmentation of the CNN to find the exact position of the crack path and crack tip. Furthermore, an extension of the method to detect fatigue crack evolution in real size structures and components will be the focus of future investigations. To this purpose, fatigue biaxial testing of skin-like fuselage-relevant structures will be investigated considering the actual load condition expected during flight of commercial aircrafts. The generation of new training data will be the main challenge to provide further generalization of the methodology.

## 4 | CONCLUSIONS

We implemented a U-Net type deep CNN to segment crack paths and their respective crack tip locations from displacement fields obtained using DIC during *fc* experiments. The CNN was trained to predict the coordinates of crack paths and crack tips based on the displacement fields around the fatigue crack. The DIC data used for training were supplemented with data from finite element calculations aiming at enlarging the amount of data. The CNN segmentation was complemented by a post-processing algorithm that helped identifying the most probable crack tip location. The performance of the trained CNN was assessed for three *fc* specimens with different geometries. The main findings of this work are as follows:

1. CNNs are an appropriate tool for detection and segmentation of fatigue cracks based on displacement fields obtained experimentally by DIC. The class imbalance must be considered during training to account for the significantly different numbers of pixels that correspond to the classes *background*, *crack path*, and *crack tip*. The background noise of the DIC system has only an influence on the performance of the neural network if the displacements are small, that is, for very low SIFs.
2. The application of the trained CNN on new data, that is, on a specimen different from the one used for training, resulted in a very reliable reproduction of the  $a$ - $N$  diagram with a small discrepancy with respect to the ground truth segmentation. Moreover, the CNN segmentation shows a prediction of the crack path and the crack tip that can be even more accurate than the manual segmentation based on images obtained from a reflex camera.
3. The addition of FEA data to the training process improves the performance of the network and increases the stability of the segmentation. This is particularly important to increase the training data for segmentation of cracks in specimens with dimensions and loading conditions different from the ones used for original CNN training.
4. The segmentation process is robust for DIC data obtained at the maximum load during the fatigue cycles owing to the larger displacement gradients around the crack. At lower nominal loads (close to zero) the inherent noise of the DIC methodology overshadows the crack tip field, which is necessary for an accurate identification of the crack tip position.

The method introduced here is a key step towards fully automatic implementation of DIC for fatigue crack growth experiments for specimens and structures with complex geometries.

## ACKNOWLEDGMENT

The authors would like to thank E. Dietrich for supporting and conducting the *fc*p experiments.

Open Access funding enabled and organized by ProjektDEAL.

## DATA AVAILABILITY STATEMENT

The data sets generated for training the CNN in the current study are available from the corresponding author on reasonable request.

## NOMENCLATURE

$a$	crack length
$a_{\text{init}}$	initial and final crack length
$a_{\text{rupture}}$	
CNN	convolutional neural network
DCPD	direct current potential drop
DIC	digital image correlation
$E$	Young's modulus
$e, e_{\text{mean}}$	root squared error of predicted and ground truth crack tip positions
$\varepsilon_{\text{VM}}$	von Mises equivalent strain
$F$	force
<i>fc</i> p	fatigue crack propagation
FEA	finite element analysis
$K, K_I$	stress intensity factor
$K_{\text{init,min}}$	stress intensity factor at initial crack length under minimum force
$K_{\text{rupture,max}}$	stress intensity factor at final rupture under maximum force
MT	middle tension
$\mu$	Poisson's ratio
$p_{\text{ct}}$	probability for class <i>crack tip</i>
$R$	force ratio
ReLU	rectified linear unit
$S_{2.0}, S_{4.7}$	specimens with thickness 2.0 and 4.7 mm, respectively
$t$	thickness
$u, u_x, u_y$	displacements
$\omega_{\text{BG}}, \omega_{\text{CP}}, \omega_{\text{CT}}$	class weights for classes <i>background</i> , <i>crack path</i> and <i>crack tip</i>

## ORCID

Tobias Strohmann  <https://orcid.org/0000-0002-9277-1376>

Eric Breitbarth  <https://orcid.org/0000-0002-3479-9143>

## REFERENCES

- Schijve J. Fatigue of structures and materials in the 20th century and the state of the art. *Int J Fatigue*. 2003;25(8):679-702.
- Swift T. Fail-Safe Design Requirements and Features, Regulatory Requirements. Paper presented at: AIAA international air and space symposium and exposition: the next 100 years 2003; Reston, Virginia.
- Schijve J. Fatigue damage in aircraft structures, not wanted, but tolerated? *Int J Fatigue*. 2009;31(6):998-1011.
- Negggers J, Allix O, Hild F, Roux S. Big data in experimental mechanics and model order reduction: today's challenges and Tomorrow's opportunities. *Arch Comput Meth Eng*. 2018;25(1):143-164.
- Mokhtarishirazabad M, Lopez-Crespo P, Moreno B, Lopez-Moreno A, Zanganeh M. Evaluation of crack-tip fields from DIC data: a parametric study. *Int J Fatigue*. 2016;89:11-19.
- Mathieu F, Hild F, Roux S. Identification of a crack propagation law by digital image correlation. *Int J Fatigue*. 2012;36(1):146-154.
- Roux S, Réthoré J, Hild F. Digital image correlation and fracture: an advanced technique for estimating stress intensity factors of 2D and 3D cracks. *J Phys D Appl Phys*. 2009;42(21):1-37.
- Becker TH, Mostafavi M, Tait RB, Marrow TJ. An approach to calculate the J-integral by digital image correlation displacement field measurement. *Fatigue Fract Eng Mater Struct*. 2012;35(10):971-984.
- Besel M, Breitbarth E. Advanced analysis of crack tip plastic zone under cyclic loading. *Int J Fatigue*. 2016;93:92-108.
- Breitbarth E, Besel M. Energy based analysis of crack tip plastic zone of AA2024-T3 under cyclic loading. *Int J Fatigue*. 2017;100:263-273.
- Lopez-Crespo P, Shterenlikht A, Patterson EA, Yates JR, Withers PJ. The stress intensity of mixed mode cracks determined by digital image correlation. *J Strain Anal Eng Des*. 2008;43(8):769-780.
- Williams ML. On the stress distribution at the base of a stationary crack. *J Appl Mech*. 1956;24(1):109-114.
- Réthoré J. Automatic crack tip detection and stress intensity factors estimation of curved cracks from digital images. *Int J Numer Methods Eng*. 2015;103(7):516-534.
- Yoneyama S, Morimoto Y, Takashi M. Automatic evaluation of mixed-mode stress intensity factors utilizing digital image correlation. *Strain*. 2006;42(1):21-29.
- Ayatollahi MR, Moazzami M. Digital image correlation method for calculating coefficients of Williams expansion in compact tension specimen. *Opt Lasers Eng*. 2017;90:26-33.
- Tumanov AV, Shlyannikov VN, Chandra Kishen JM. An automatic algorithm for mixed mode crack growth rate based on drop potential method. *Int J Fatigue*. 2015;81:227-237.
- Tarnowski KM, Nikbin KM, Dean DW, Davies CM. A unified potential drop calibration function for common crack growth specimens. *Exp Mech*. 2018;58(6):1003-1013.
- Smith RF, Doig P. Crack-length measurement by compliance in fracture-toughness testing. *Exp Mech*. 1986;26(2):122-127.
- Sullivan AM, Crooker TW. A crack-opening-displacement technique for crack length measurement in fatigue crack growth rate testing—development and evaluation. *Eng Fract Mech*. 1977;9(1):159-166.
- Khalili K, Vahidnia M. Improving the accuracy of crack length measurement using machine vision. *Procedia Technol*. 2015;19:48-55.
- Simunek D, Leitner M, Grün F. In-situ crack propagation measurement of high-strength steels including overload effects. *Procedia Eng*. 2018;213:335-345.

22. Zhao J, Sang Y, Duan F. The state of the art of two-dimensional digital image correlation computational method. *Eng Reports*. 2019;1(2):1-15.
23. Gao Z, Xu X, Su Y, Zhang Q. Experimental analysis of image noise and interpolation bias in digital image correlation. *Opt Lasers Eng*. 2016;81:46-53.
24. Cha Y-J, Choi W, Büyükköztürk O. Deep learning-based crack damage detection using convolutional neural networks. *Comput Aided Civ Inf Eng*. 2017;32(5):361-378.
25. Li S, Zhao X. Image-based concrete crack detection using convolutional neural network and exhaustive search technique. *Adv Civil Eng*. 2019;2019(2):1-12.
26. Mandal V, Uong L, Adu-Gyamfi Y. Automated Road Crack Detection Using Deep Convolutional Neural Networks. Paper presented at: 2018 IEEE international conference on big data (big data); 2018.
27. Fan R, Bocus MJ, Zhu Y, et al. Road Crack Detection Using Deep Convolutional Neural Network and Adaptive Thresholding. Paper presented at: 2019 IEEE Intelligent Vehicles Symposium (IV); 2019:474-479.
28. Mohamed YS, Shehata HM, Abdellatif M, Awad TH. Steel crack depth estimation based on 2D images using artificial neural networks. *Alex Eng J*. 2019;58(4):1167-1174.
29. Dung CV, Sekiya H, Hirano S, Okatani T, Miki C. A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks. *Automation in Construction*. 2019;102:217-229.
30. ASTM. *ASTM E647 Standard Test Method for Measurement of Fatigue Crack Growth Rates*. 2013.
31. Schwalbe K-H, Hellmann D. Application of the electrical potential method to crack length measurements using Johnson's formula. *J Test Eval*. 1981;218-220.
32. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*; 2015:234-241.
33. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(4):640-651.
34. Liu W, Rabinovich A, Berg A. ParseNet: looking wider to see better. 2015. ArXiv:1506.04579:1-11.
35. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. 2018. ArXiv:1703.06870v3:1-12.
36. Liu Z, Cao Y, Wang Y, Wang W. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Autom Constr*. 2019;104:129-139.
37. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. *Z Med Phys*. 2019; 29(2):86-101.
38. Maas AL, Hannun AY, Ng AY. Rectifier Nonlinearities Improve Neural Network Acoustic Models. Paper presented at: Proceedings of the 30th International Conference on Machine Learning 2013; Atlanta, Georgia, USA.
39. PyTorch. <https://pytorch.org/>
40. Kingma DP, Lei BJ. Adam: a method for stochastic optimization. In: *Proceedings in ICLR*; 2015.
41. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data*. 2019;6(27):1-54.
42. Breitbarth E, Strohmann T, Requena G. High stress fatigue crack propagation in thin AA2024-T3 sheet material. *Fatigue Fract Eng Mater Struct*. 2020;43(11):2683-2693.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Strohmann T, Starostin-Penner D, Breitbarth E, Requena G. Automatic detection of fatigue crack paths using digital image correlation and convolutional neural networks. *Fatigue Fract Eng Mater Struct*. 2021;44:1336–1348. <https://doi.org/10.1111/ffe.13433>